K. Biemann
M.I.T., Cambridge, Mass.

Title

(Technical report on the progress of work sponsored by Grant NsG-211-62 of the National Aeronautics and Space Administration (covering the period from August 16, 1964 to June 15, 1965).)

Investigations conducted during the period stated above have been concerned with three areas:

I. Continuation of the collecting of mass spectra of compounds of biological interest.

II. Development of mass spectrometric techniques for the detection and identification of traces of organic compounds in geological or extraterrestrial specimens.

III. Computer techniques for the processing of high resolution mass spectral data and the computer-aided interpretation of mass spectra.

Ad I: Our earlier work (see previous progress reports) on the determination and interpretation of the mass spectra of organic compounds of biological interest has been confirmed by reinvestigations of many representative compounds using a high resolution mass spectrometer. Such an instrument permits one to determine the mass of any ion with an accuracy of a few parts in one million, i.e., with an error of only about one millimass unit. From such data one can then deduce the number and kind of elements present, i.e., the elemental composition. The tremendous value of such measurements becomes obvious if one keeps in mind that they can be made on as little as a microgram of material or even less.

As an example a few of the data obtained with uridine should be mentioned. Its conventional mass spectrum had been obtained earlier[1] and was shown in the last progress report. Accurate mass measurements on all the ions revealed the correctness of our earlier

interpretation but also shows how valuable this kind of information were if one would not know anything about the nature of the compound.

The molecular ion was found to have a mass of 244.0699 indicating the elemental composition $C_9H_{12}N_2O_6$ (calcd.: 244.0695). Among the many fragment ions were two of considerable abundance, which had masses of 113.0358 and 133.0509 which correspond to $C_4H_5N_2O_2$ (calcd.: 113.0351) and $C_5H_9O_4$ (calcd.: 133.0500) respectively. Such a result requires that there are two widely different portions in the molecule, one oxygen-rich and nitrogen-free ($C_5H_9O_4$), the other contains two nitrogens and two oxygens ($C_4H_5N_2O_2$), together containing all the carbon atoms of the molecule. Obviously, these two parts represent the ribose portion, and the uracil ring of uridine, a conclusion very important for the correct interpretation of the spectrum if the nature of the sample would not have been known. Similar conclusions are reached when considering the elemental compositions of all other ions in the spectrum. The values also are in perfect agreement with the earlier predictions[1] based on the conventional mass spectra of nucleosides.

Similar results have been obtained with amino acids, peptides, purines, carbohydrates, other nucleosides, fatty acids and steroids.

Ad II. Since our previous work had established that one can use mass spectra to identify organic compounds present in the microgram to nanogram range, we turned to the more difficult problem of doing this with natural mixtures as they are possibly present in geological specimens or meteorites.

Having realized the much greater potential of high resolution mass spectra over conventional ones, we used the latter only to work out experimental techniques, but collected the actual data with the

double focusing, high resolution instrument.

In order to avoid any contamination of the sample or the creation of artifacts, we explored direct evaporization of the organic material out of the inorganic matrix into the ion source of the spectrometer at the very low pressure that exists in the spectrometer. This has the advantages of reducing the handling of the sample to an absolute minimum (grinding and transfer into a glass vial) and may, incidentally, be the method of choice when the sample has to be taken and the spectrum determined automatically as would be the case by a soft handling instrument on another planet.

Using as little as 10-50 mg. of ground material (sediment or meteorite) and heating it slowly while continuously taking the spectra of the vapor over the specimen one obtains a very complex set of data (for each spectrum a few hundred of accurate masses corresponding to the ions formed from the organic compounds present) which is converted by a computer to an interpretable form, namely an array of the elemental combinations of all the ions produced, which thus reflects the types of compounds present.

As an example, Figure 1 represents part of the volatiles obtained from a sample of cretaceous sediment. It is clearly obvious that there are in fact many hydrocarbons evolved (C,H-column) which range in saturation from highly saturated ones (high H to C ratio) to highly unsaturated ones (low H/C ratio), the asterisks indicating the abundance of the various species on a logarithmic scale. But there are also present many oxygenated species, mainly aromatic ones (C,H,O-column) whose C/H ratio indicates substituted benzofurans, while the scarceness of nitrogen containing ions is striking. Considering that this is only a small part of the information obtained

with as little as 50 mg. of the sediment (containing only a very small amount of organic material), it is a wealth of specific data. A more detailed discussion of this type of mass spectra and their interpretation has been published[2].

We have more recently employed this technique in an effort to "map" the interior of a piece of meteorite (Murray), by drilling a narrow hole (2-3 mm I.D.) through the specimen, collecting the material corresponding to 3 mm deep sections (a total of eight) and recording the mass spectra of the organic constituents in the manner outlined above. It is hoped that such a technique would enable one to obtain a permanent, very detailed record of the general distribution of organic material in a meteorite, to have some information concerning their nature and to deduce whether they may possibly be artifacts in which case they would be present on and near the surface, but not in the center.

Preliminary interpretation of the spectra of the eight Murray samples mentioned above indicate that the hydrocarbons are distributed rather equally throughout the mateorite specimen while the oxygenated compounds seem to slightly predominate on the surface.

The detailed interpretation of the data is somewhat hampered by the relative slowness of the measurement of the accurate position of the many hundred lines produced by the spectrometer on a photographic plate. While the actual recording of these spectra takes only a few seconds for each, measuring all the lines of the spectrum with a reliable accuracy of $\pm$ 0.5 micron takes from 2-6 hours with our present comparator. It is hoped that we can acquire a fully automatic comparator with IBM compatible digital readout on magnetic tape, to overcome this data handling problem.

Ad III. During this work it became apparent that not only does the processing of these data require a computer because it could not conceivably be done by hand (each line position is a seven digit number, has to be converted to the corresponding mass with the same accuracy and must then be correlated with that combination of carbon, hydrogen, nitrogen, oxygen, sulfur atoms, etc., which add up to exactly the same mass) but also that these data lend themselves exceedingly well to computer interpretation. In contrast to conventional mass spectra, which yield only the integral mass of the ions and may thus correspond to very many possible elemental compositions, the high resolution data are much more directly and uniquely related to the nature of the compound because one directly determines the number and kind of elements that make up the compound itself as well as all the fragments produced on electron impact.

Thus we are presently developing programs which will enable the computer to pre-interpret the data in terms of the general nature of the compounds giving rise to the particular spectrum, the functional groups present, their location, etc. Eventually, the computer may be able to arrive at the complete answer, namely the correct structure of the compound.

As a first step we have reached the point where the computer can uniquely determine the molecular weight and elemental composition of the compound in question, even in the presence of impurities or if the mass spectrum should lack the peak due to the molecular ion if one deals with a compound that fragments with extreme ease. Both cases are frequent reasons for the misinterpretation of mass spectra.

The principle of the technique is based on the fact that the molecule must have all valences occupied, cannot contain heavy

isotopes and must be reconcilable with all other ions in the spectrum (except those due to impurities and contaminants) which must all be derived from the actual molecule by the loss of small groups of atoms. By having the computer test a number of peaks at the high mass end of the spectrum for these characteristics, it selects the one that fits best and thus must be the true molecular ion. It also tests whether there is another still more compatible combination of elements which can be "synthesized" by adding the weight of the small groups of atoms to the existing ions of high mass. This is based on the fact that if a molecule fragments very easily, it does so by the loss of discrete groups of atoms.

The program has been tested successfully with many compounds. For example, when the high resolution spectrum of androsterone acetate was run with this program, the computer selected the correct molecular ion, $C_{21}H_{32}O_3$. Even when the peak due to this ion was removed from the set of data, the same, correct result was obtained because the computer could reconcile the remaining data only by assuming that the molecular composition of the compound is $C_{21}H_{32}O_3$, in spite of the absence of an ion corresponding to this composition. This technique was the subject of a recent publication[3] of which reprints are enclosed.

This is only the beginning of what we call "computer-aided interpretation of mass spectra". Further insight into the nature of the compound can be obtained from a consideration of the elemental composition of the fragments that are easily lost from the molecule. Thus in the case of androsterone acetate, these are CH_3, $CH_2{=}CH_2$, CO, CH_3COOH and various combinations thereof, which allow the conclusion that it is an O-acetate, contains a cyclic carbonyl

group, one or more methyl groups, no alkyl side chain and a total of six double bonds or rings. From these facts it would not be difficult to arrive at the conclusion that one deals with an acetoxyketo-steroid with a total of 19 carbon atoms in the carbon skeleton, if the structure of the compound were not known.

We are presently working on the elaboration of computer programs that would carry out such interpretations efficiently and logically. The implication of such techniques to the investigation of organic materials by remotely controlled mass spectrometers are of obvious significance.

References:

(1) Application of Mass Spectrometry to Structure Problems. VI. Nucleosides. K. Biemann and J.Am McCloskey, J. Am. Chem. Soc., 84, 2005 (1962).

(2) a. Element-Mapping, A New Approach to the Interpretation of High Resolution Mass Spectra, K. Biemann, P. Bommer and D.M. Desiderio, Tetrahedron Letters, No. 26, 1725 (1964). b. High Resolution Mass Spectrometry of Natural Products, K. Biemann, J. Pure Appl. Chem., 9, 95 (1964). c. New Techniques for the Interpretation of High Resolution Mass Spectra of Organic Molecules, K. Biemann, P. Bommer, D.M. Desiderio and W.J. McMurray. Presented at Symposium on Mass Spectrometry, Paris, September 1964. (To be published in "Advances in Mass Spectrometry," Vol. III.)

(3) Computer-Aided Interpretation of High Resolution Mass Spectra, K. Biemann and W.J. McMurray, Tetrahedron Letters, No. 11, 647 (1965).

figure 3: ELEMENT MAP JUDY CREEK CRETACEOUS SEDIMENT, 300 243-07-1

The structures and identifications noted are all tentative.

error, mass found - mass calculated
millimass units

relative intensity

Number of C atoms

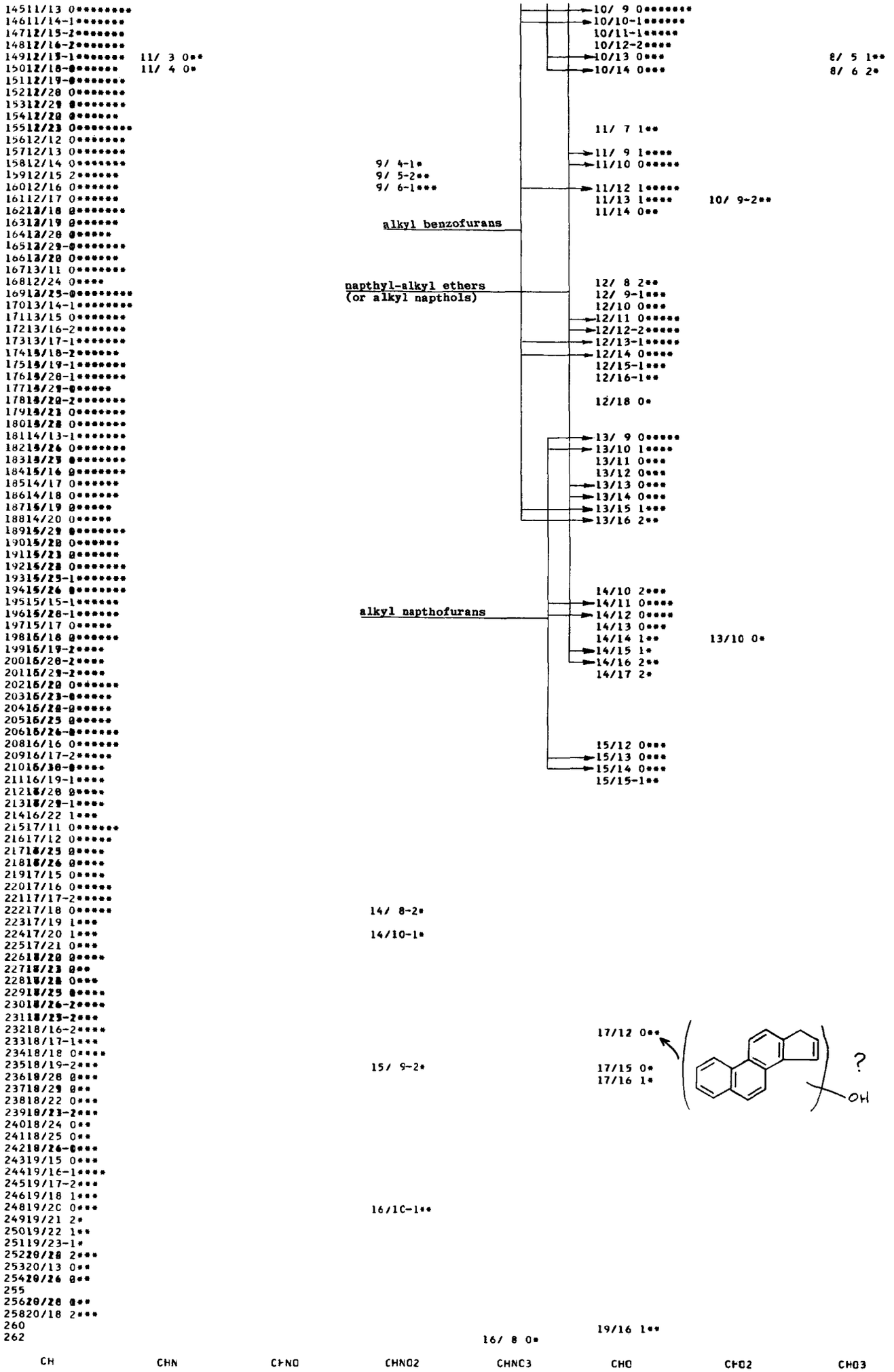
Number of H atoms

an overprint due to two CH ions of the same nominal mass. In this case, 6/14 and 7/2

alkyl furans

phenyl-alkyl ethers (or alkyl phenols)

Nominal Mass	CH	CHN	CHNO	CHNO2	CHNO3	CHO	CHO2	CHO3
32								
33								
34								
36	3/ 0 0							
37	3/ 1 0							
38	3/ 2 0							
39	3/ 3-1							
40	3/ 4-1							
41	3/ 5-1					2/ 1 0		
42	3/ 6-1					2/ 2 0		
43	3/ 7 0		1/ 1 0			2/ 3 0		
44	3/ 8-1					2/ 4 0	1/ 0 0	
45							1/ 1-1	
46				0/ 0 0				
47								
48								
49	4/ 1 0							0/ 1-1
50	4/ 2 0							
51	4/ 3 0							
52	4/ 4 0	3/ 2 0						
53	4/ 5 0					3/ 1 0		
54	4/ 6-1	3/ 4 0				3/ 2 0		
55	4/ 7 0					3/ 3 0		
56	4/ 8 0		2/ 2-1			3/ 4 0		
57	4/ 9 0		2/ 3 2			3/ 5 0		
58	4/10-2					3/ 6 0		
59						3/ 7 0	2/ 3 1	
60							2/ 4 0	
61	5/ 1 1							1/ 1 0
62	5/ 2 1							
63	5/ 3 0							
64	5/ 4-1							
65								
66	5/ 6-2							
67	5/ 7-2		3/ 1 1					
68	5/ 8-2					4/ 4-1		
69	5/ 9-1		3/ 3 0			4/ 5-1		
70	5/10 0		3/ 4 0			4/ 6-1		
71	5/11 0		3/ 5 1			4/ 7 0		
72			3/ 6 1			4/ 8-1		
73	6/ 1 1							
74	6/ 2 0							
75	6/ 3 0							
76	6/ 4 0							
77	6/ 5 1							
78	6/ 6 0							
79	6/ 7 0							
80	6/ 8 0							
81	6/ 9 0					5/ 4-1		
82	6/10 0					5/ 5 0		
83	6/11 0		4/ 5 2			5/ 6 0		
84	6/12 0		4/ 6 2			5/ 7 0		
85	6/13 0		4/ 7 2			5/ 8 0	4/ 4 2	
86	6/14-2					5/ 9 0		
87	7/ 3 1					5/10 1		3/ 2 0
88	7/ 4 1							
89	7/ 5 0							
90	7/ 6 0							
91	7/ 7 0						4/11-1	
92	7/ 8 0							
93	7/ 9 0							
94	7/10 0	6/ 8 1				6/ 5 0		
95	7/11 0					6/ 6 0		
96	7/12-1					6/ 7 0		
97	7/13 0		5/ 6 2			6/ 8 0	5/ 4 0	
98	7/14-1		5/ 7 0			6/ 9 0		4/ 1 0
99	7/15 0					6/10 0		4/ 2-2
100	7/16 0					6/11 0		
101	8/ 5 1					6/12 1		4/ 4-2
102	8/ 6 0							
103	8/ 7 0					7/ 2-2		
104	8/ 8 0	7/ 6 2						
105	8/ 9 0					7/ 5 0	6/ 1-2	
106	8/10-1					7/ 6-1	6/ 2-2	
107	8/11 0					7/ 7 0		
108	8/12-1					7/ 8-1		
109	8/13 0					7/ 9 0		
110	8/14 0		6/ 8 2			7/10 0	6/ 6 0	5/ 2 0
111	8/15 0					7/11 0		5/ 3-2
112	8/16 0	8/ 2 1				7/12 0		
113	9/15 0							5/ 5-1
114	9/16 0	8/ 4 0						5/ 6-2
115	9/ 7 1						6/11 0	
116	9/ 8-1							
117	9/ 9 0							
118	9/10-1					8/ 6 0		
119	9/11 0					8/ 7 0		
120	9/12-2					8/ 8 0		
121	9/13 0					8/ 9 0		
122	9/14 0					8/10 0		
123	9/15 0					8/11 0	7/ 7 1	
124	10/16 0					8/12 0		
125	10/15 0	9/ 3 0						
126	10/18 0	9/ 4-2				8/14 2		6/ 6 0
127	10/19 0							
128	10/20 0							
129	10/ 9-2							
130	10/10-2							
131	10/11-1					9/ 7 0	7/15 2	
132	10/12-1					9/ 8 0		
133	10/13-1					9/ 9 0		
134	10/14 0					9/10 0		
135	10/15-2					9/11 0		
136	10/16-1					9/12 0		
137	10/15 0					9/13-1		
138	10/18-1					9/14 0		
139	10/19-0							
140	10/20-1							
141	10/21 0					9/17-1		
142	10/20-0							
143	11/11-2							
144	11/12-1					10/ 8 0		


alkyl benzofurans
napthyl-alkyl ethers
(or alkyl napthols)
alkyl napthofurans
11/ 3 0**
11/ 4 0*
9/ 4-1*
9/ 5-2**
9/ 6-1***
10/ 9 0*******
10/10-1******
10/11-1*****
10/12-2****
10/13 0***
10/14 0***
8/ 5 1**
8/ 6 2*
11/ 7 1**
11/ 9 1****
11/10 0*****
11/12 1*****
11/13 1****
11/14 0**
10/ 9-2**
12/ 8 2**
12/ 9-1***
12/10 0***
12/11 0*****
12/12-2*****
12/13-1*****
12/14 0****
12/15-1***
12/16-1**
12/18 0*
13/ 9 0*****
13/10 1****
13/11 0***
13/12 0***
13/13 0***
13/14 0***
13/15 1***
13/16 2**
14/10 2***
14/11 0****
14/12 0****
14/13 0***
14/14 1**
14/15 1*
14/16 2**
14/17 2*
13/10 0*
15/12 0***
15/13 0***
15/14 0***
15/15-1**
14/ 8-2*
14/10-1*
15/ 9-2*
16/1C-1**
16/ 8 0*
17/12 0**
17/15 0*
17/16 1*
19/16 1**
?
OH
CH
CHN
CHNO
CHNO2
CHNO3
CHO
CHO2
CHO3